УДК 004.274

И.Г. Герасимов, С.В. Терещенко
Институт проблем искусственного интеллекта МОН Украины и НАН Украины, г. Донецк
Украина, 83048, г. Донецк, ул. Артёма, 118б

Построение ER-модели молекулярного состава клетки как основа функционирования биокомпьютера (цитокомпьютера)

I.G. Gerasimov, S.V. Tereshchenko
Institute of Artificial Intelligence MES Ukraine and NAS Ukraine, Donetsk
Ukraine, 83048, Donetsk, Artema st., 118b

Construction of ER-Model of the Molecular Composition of Cell as the Basis on Functioning of Biocomputer (Cytocomputer)

І.Г. Герасимов, С.В. Терещенко
Інститут проблем штучного інтелекту МОН України і НАН України, м. Донецьк
Україна, 83048, м. Донецьк, вул. Артема, 118б

Побудова ER-моделі молекулярного составу клітини як основа функціонування біокомп'ютера (цітокомп'ютера)

В работе обсуждается возможность разработки цитокомпьютера – биокомпьютера, основанного на алгоритмах клеточного функционирования и имеющейся материальной базе (микросхемы). С этой целью предпринята первая попытка построения модели «сущность-связь» (entity-relationship model, ER-model) молекулярной структуры клетки и определены соответствующие понятия. Разработанная ER-модель молекулярного состава клетки является основой для создания цитокомпьютера.
Ключевые слова: биокомпьютер, цитокомпьютер, клетка, молекулярная структура клетки, модель «сущность-связь», алгоритм функционирования клетки.

The paper discusses the possibility for developing of a cytocomputer. Cytocomputer is a biocomputer which based on algorithms of cellular functioning and on available material base (chips). For this purpose was made the first attempt to construct an ER-model of the molecular structure of cells and were defined the relevant concepts. Developed ER-model of the molecular structure of a cell is the basis for creating of the cytocomputer.
Key words: biocomputer, cytocomputer, cell, molecular structure of cell, entity-relationship model, algorithm of cellular functioning.

У роботі обговорюється можливість розробки цітокомпьютера – біокомп'ютера, заснованого на алгоритмах клітинного функціонування та наявної матеріальної бази (мікросхеми). З цією метою здійснено першу спробу побудови моделі «сутність-зв'язок» (entity-relationship model, ER-model) молекулярної структури клітини та визначено відповідні поняття. Розроблена ER-модель молекулярного складу клітини є основою для створення цітокомпьютера.
Ключові слова: біокомп'ютер, цітокомп'ютер, клітина, молекулярна структура клітини, модель «сутність-зв'язок», алгоритм функціонування клітини.

Полупроводниковые технологии, посредством которых стабильно развивается современная компьютерная индустрия, в настоящее время быстро приближаются к своим физическим, идеологическим и конструктивным пределам. Развитие компьютеров с использованием технологии кремниевых микрочипов может быть ограничено вследствие невозможности их бесконечной дальнейшей миниатюризации и ростом объемов обмена информацией между центральным процессором и оперативной памятью. Порогом производительности кремниевых микрочипов (микросхем) является выполнение 1018 операций с плавающей запятой в секунду [1]. По-видимому, невозможно преодолеть этот рубеж с помощью традиционных дискретных операций с плавающей запятой. В связи с этим возникает необходимость создания компьютеров, функционирование которых коренным образом отличается от указанной технологии.

Альтернативой классических компьютеров могут стать биокомпьютеры. Биокомпьютер – это компьютер, который функционирует как живой организм либо содержит биологические компоненты. В последнем случае в целях создания биокомпьютеров используют такое направление исследований, как молекулярные вычисления. В качестве вычислительных элементов используют белки и нуклеиновые кислоты. Идея о таких вычислительных элементах находится в стадии развития. Проводятся как in vitro, так и in vivo исследования (например, [8-10]). Многообещающие результаты этих исследований позволяют надеяться на большие достижения в этом направлении [3].

Другой подход к созданию биокомпьютера основан на использовании классической материальной базы (кремниевых микрочипов) и адаптивных алгоритмов, аналогичных тем, которые лежат в основе функционирования живых систем, что позволит существенно уменьшить число операций с плавающей запятой. Элементарной живой системой, которая может функционировать как самостоятельно, так и в составе многоклеточного организма, безусловно, является клетка. Клетка содержит множество низко- и высокомолекулярных соединений, находящихся между собой в постоянном взаимодействии, что определяет ее жизнедеятельность. Следовательно, прежде чем перейти к разработке соответствующих алгоритмов для создания биокомпьютеров такого рода (цитокомпьютера), необходимо рассмотреть во взаимосвязи молекулярный состав, а затем и надмолекулярную структуру клетки и моделировать их. Решению первой задачи, а именно моделированию молекулярной структуры клетки и посвящена настоящая работа.

В клетке вычислительные процессы осуществляются с помощью межклеточных сигналов, реакция на которые заложена в программу клеточного функционирования. В межклеточном обмене информацией сигналом называется действие, когда клетка активирует определенные молекулы. Существует два способа такой активации, приводящих в обоих случаях к изменению клеточного микроокружения: 1) молекула (точнее, ее часть) оказывается на внешней стороне клеточной мембраны; 2) молекула выделяется клеткой во внеклеточную (межклеточную) среду. В зависимости от вида полученного сигнала ответной реакцией клетки является продуцирование молекул определенного вещества либо изменение состояния, поведения клетки и т.д. На каждый тип молекул адекватно откликаются только те клетки, в программу которых заложена идентификация данного конкретного сигнала.

Функционирование и поведение живой клетки, в первую очередь, обусловливает множество содержащихся в ней нуклеиновых кислот (дезоксирибонуклеиновой –

ДНК и рибонуклеиновой – РНК) и белков, которые находятся в динамическом взаимодействии друг с другом, а также множество других молекул. Обратимые белок-белковые взаимодействия представляют собой один из видов динамических процессов. В настоящее время становится признанным, что большинство белков эукариот являются мультимодульными и полифункциональными [4]. Каждый из модулей может выполнять самостоятельную функцию, в результате чего белок приобретает способность выполнять целый комплекс различных функций.

Благодаря мультимодульности и полифункциональности большинства белков эукариот, их сложные белковые сети могут переплетаться друг с другом. Способность белков восстанавливать правильную трёхмерную структуру после денатурации позволила выдвинуть гипотезу о том, что вся информация о конечной структуре белка содержится в его аминокислотной последовательности [5]. Аналогичным, а также иным образом и другие молекулы, содержащиеся в клетке, участвуют в управлении ее функционированием.

Прежде, чем приступить к созданию системы автоматизированной обработки информации, основанной на алгоритмах клеточного функционирования (иначе говоря, к созданию биокомпьютера, или цитокомпьютера) или хотя к моделированию подходящих систем, необходимо сформировать понятия о предметах, фактах и событиях, которыми будет оперировать данная система. Для того чтобы привести эти понятия к той или иной модели данных, необходимо заменить их информационными представлениями. Одним из наиболее удобных инструментов унифицированного представления данных, независимо от реализующего его программного обеспечения, является модель «сущность-связь» (entity-relationship model, ER-model) [6].

Модель «сущность-связь» (ER-модель) основана на некой важной семантической информации о реальном мире и предназначена для логического представления данных. В ней определены значения данных в контексте их взаимосвязи с другими данными. Важным является тот факт, что на основании модели «сущность-связь» могут быть порождены все существующие модели данных (иерархическая, сетевая, реляционная, объектная), поэтому она является наиболее общей. В рамках такой модели любой фрагмент предметной области может быть представлен как множество сущностей, между которыми существует некоторое множество связей.

Для построения ER-модели молекулярной структуры клетки может быть использован некий гибрид нотаций Чена (обозначение сущностей, связей и атрибутов) и Мартина (обозначение степеней связей) [7]. В таблице приведен список соответствующих обозначений.

Процесс построения ER-модели состоит из нескольких этапов, среди которых:

- идентификация представляющих интерес сущностей и связей;
- идентификация семантической информации в наборах сущностей и связей;
- определение атрибутов сущностей и наборов их значений;
- определение степеней связей;
- организация данных в виде отношений «сущность-связь».

Для построения ER-модели молекулярного состава клетки выделим интересующие сущности и связи. Клетка состоит из множества молекул неорганических (воды, минеральных солей, неорганических фосфатов, катионов металлов и пр.) и органических (белков, углеводов, жиров, нуклеиновых кислот, органических фосфатов и пр.) соединений, которые в ER-модели представляются ее сущностями.

Таблица 1 – Условные обозначения ER-модели

Обозначение	Содержание
Имя сущности	Сущность (объект системы)
… …	Группа сущностей (сущности с одинаковым набором атрибутов)
Имя атрибута	Атрибут сущности
Имя атрибута	Ключевой атрибут сущности
Имя набора атрибутов … …	Набор атрибутов сущности или группы сущностей
Имя связи	Набор связей между сущностями
	Связь между атрибутом и сущностью
	Связь между сущностями
	Степень связи «один-ко-многим»
	Модальность связи «может»
	Модальность связи «должен»

Если сущности имеют одинаковые атрибуты, то они группируются в набор сущностей.

Далее представлены сущности, наборы сущностей, а также их атрибуты, использованные для моделирования молекулярного состава клетки.

1. Сущность {«Клетка»} представляет собой наиболее высокий уровень сущности в данной модели. Набором ее атрибутов являются «Наименование», «Класс», «Тип», «Состояние», «Условие изменения состояния». Обозначим их как набор атрибутов 1 (рис. 2).

2. Сущность {«Вода»} характеризуется таким набором атрибутов «Концентрация в клетке» и «Компартментализация». Обозначим их как набор атрибутов 2 (рис. 2).

3. Набор сущностей {«Газ», «Минеральная соль», «Фосфат неорганический», «Катион металла»} включает набор атрибутов 2 и атрибуты «Наименование», «Класс», «Структура», «Процентное содержание в клетке», «Степень растворимости», «Функции». Обозначим его как набор атрибутов 3 (рис. 3).

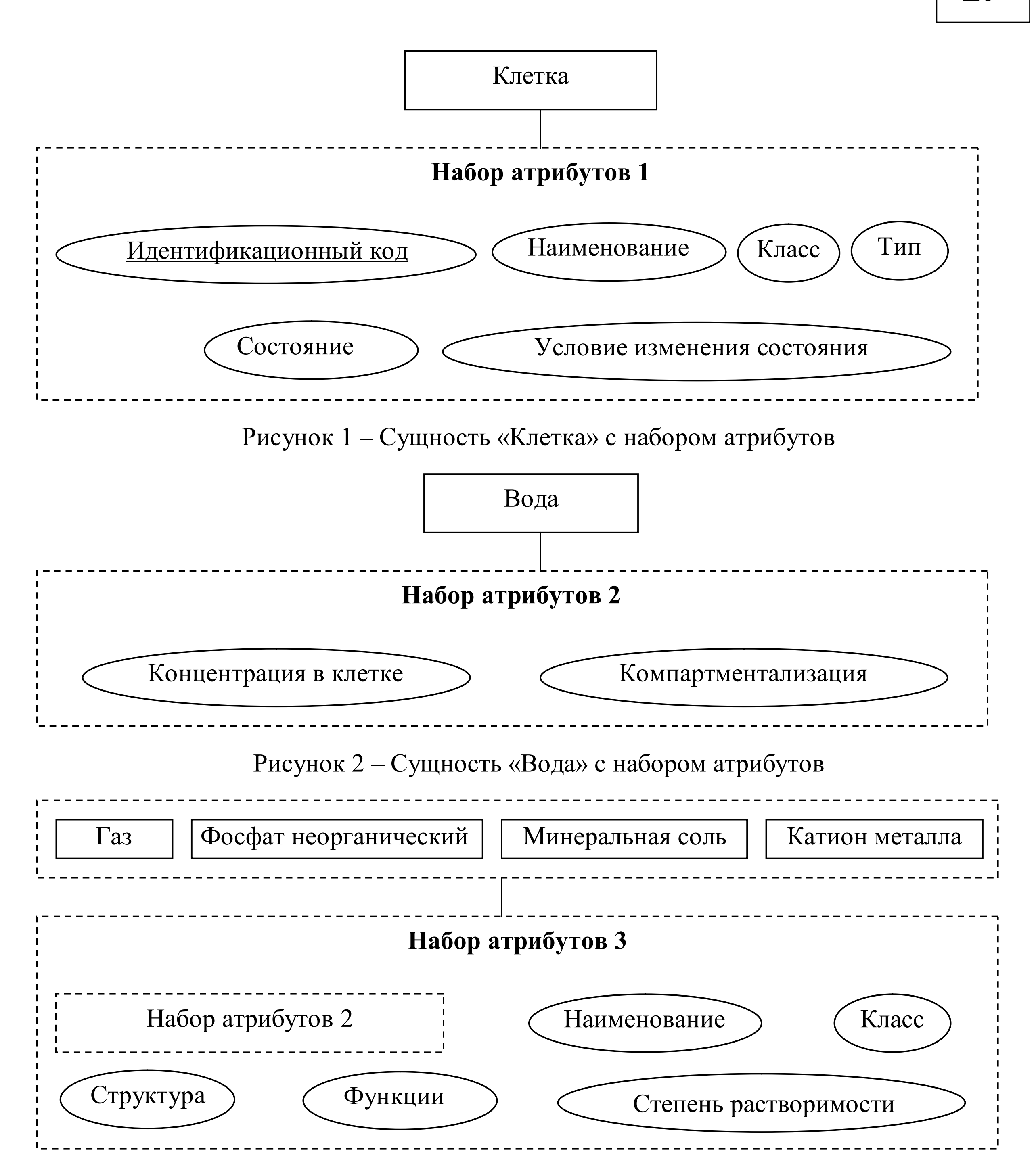


Рисунок 1 – Сущность «Клетка» с набором атрибутов

Рисунок 2 – Сущность «Вода» с набором атрибутов

Рисунок 3 – Группа сущностей «Газ», «Минеральная соль», «Фосфат неорганический», «Катион металла» с атрибутами

4. Набор сущностей {«Нуклеиновая кислота», «Белок», «Фосфат органический», «Углевод»} включает в себя набор атрибутов 3 и атрибут «Молекулярная масса». Обозначим его как набор атрибутов 4 (рис. 4).

5. Сущность «Жир» имеет набор атрибутов 4. От предыдущей группы сущностей его отличает только то, что его наличие в клетке необязательно (рис. 5).

6. В группу сущностей {Другое} выделено все множество других элементов, которое может содержаться в клетке и в данной модели эта сущность не атрибутирована.

Каждая связь имеет два конца и одно или два наименования. Наименование обычно выражается в неопределенной глагольной форме: «иметь», «принадлежать» и т.п. Каждое из наименований относится к своему концу связи. Иногда наименования не записываются ввиду их очевидности.

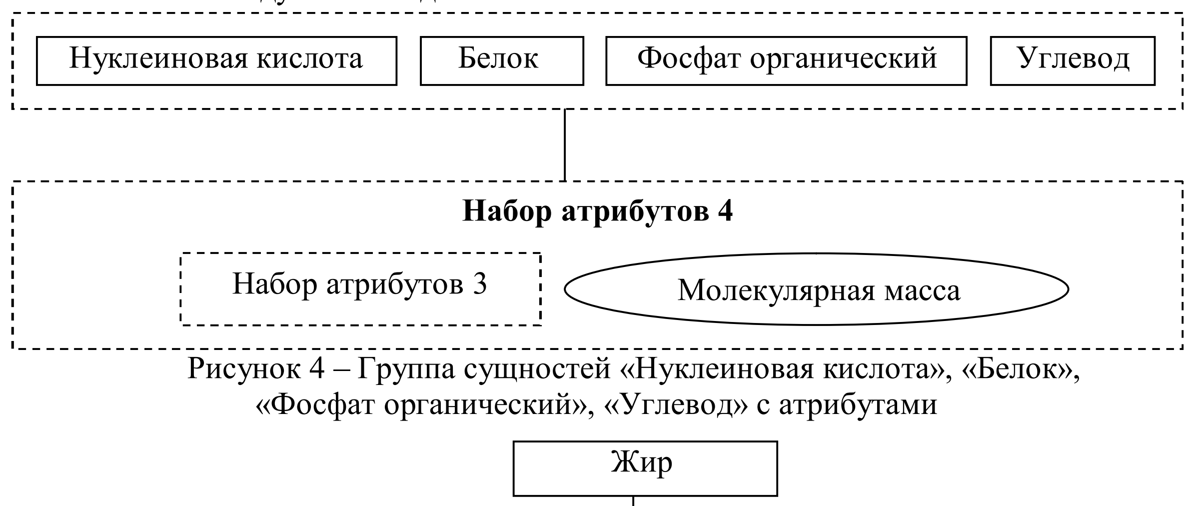


Рисунок 4 – Группа сущностей «Нуклеиновая кислота», «Белок», «Фосфат органический», «Углевод» с атрибутами

Жир

Набор атрибутов 4

Рисунок 5 – Сущность «Жир» с набором атрибутов

Каждая связь имеет степень (о чем далее) и быть одной из следующих типов: «один-к-одному», «один-ко-многим», «много-ко-многим» (табл. 1). Связь типа «один-к-одному» означает, что один экземпляр одной сущности связан с одним экземпляром другой сущности и чаще всего свидетельствует о том, что на самом деле есть всего одна сущность, неправильно разделенная на две. Связь типа «один-ко-многим» означает, что один экземпляр одной сущности связан с несколькими экземплярами другой сущности. Это наиболее часто встречающийся тип связи. Сущность со стороны «один» называется родительской, а сущность со стороны «много» – дочерней. Связь типа «много-ко-многим» означает, что каждый экземпляр одной сущности может быть связан с несколькими экземплярами другой сущности, и, наоборот. Такой тип связи является временным типом связи, допустимым на ранних этапах разработки модели и в дальнейшем должен быть заменен двумя связями типа «один-ко-многим» путем создания промежуточной сущности.

Приведем примеры связей разных типов. Пример связи «один-к-одному»: у одного нейрона может быть только один аксон. Пример связи «один-ко-многим»: на одном дереве может быть много плодов. Пример связи «многие-ко-многим»: одна пчела может опылить много разных цветков, но в то же время один и тот же цветок могут опылить множество разных пчел.

Степень связи показывает, какое максимальное количество экземпляров данной сущности может участвовать в конкретной связи, точнее, в одном экземпляре связи (табл. 1). Поскольку рассматривается бинарная связь, то ее характеризуют две степени связи для обеих сущностей, участвующих в ее образовании. Для связи «один-к-одному» степени равны 1 и 1, для связей «один-ко-многим» – 1 и N; для связи «многие-к-одному » – N и 1 (N – количество экземпляров сущностей одного вида, связанных с одним экземпляром сущности другого вида); для связи «многие-ко-многим» – N и M (N – количество экземпляров сущности одного вида и M – количество экземпляров сущности другого вида, связанные между собой).

Каждая связь может иметь одну из двух модальностей связи (табл. 1). Модальность «может» означает, что экземпляр одной сущности может быть связан с одним или несколькими экземплярами другой сущности, а может быть и не связан ни с одним экземпляром. Модальность «должен» означает, что экземпляр одной сущности обязан быть связан не менее чем с одним экземпляром другой сущности. Модальности разных концов связи могут быть разными.

Поскольку в одной клетке множество молекул различных веществ тем или иным образом взаимодействуют с другими, постольку необходимо определить связи вида «один-ко-многим» с наименованием «содержит».

Описанный графический синтаксис позволяет однозначно читать диаграммы ER-модели (ER-диаграммы), пользуясь следующей схемой построения фраз: <Каждый экземпляр СУЩНОСТИ 1> <МОДАЛЬНОСТЬ СВЯЗИ> <НАИМЕНОВАНИЕ СВЯЗИ> <ТИП СВЯЗИ> <экземпляр СУЩНОСТИ 2>. Каждая связь может быть прочитана как слева направо, так и справа налево [7]. Например, диаграмму, изображенную на рис. 6, следует читать следующим образом слева направо <Каждая КЛЕТКА> <ДОЛЖНА> <СОДЕРЖАТЬ><один или несколько><экземпляров МОЛЕКУЛЫ ВОДЫ>, или справа налево: <Молекула воды><Может><Содержаться в><Клетке>.



Рисунок 6 – Пример ER-диаграммы

На рис. 7 представлена модель «Сущность-связь» молекулярного состава клетки. Разработанная ER-модель описывает молекулярный состав клетки и формализует связи между различными структурными компонентами клетки. На ее основе представляется возможным в дальнейшем моделировать функционирования живой клетки, используя соответствующие алгоритмы и наличную материальную базу (микросхемы).

Развиваемый новый подход в развитии биокомпьютерных технологий, основанный на принципах функционирования клетки как отдельного электронного устройства, является перспективным. Разработана ER-модель молекулярного состава клетки, которая позволяет в дальнейшем перейти к созданию биокомпьютера (цитокомпьютера). Однако данная модель не отображает внутриклеточную структуру на надмолекулярном уровне, для чего необходимо выделить новый, более высокий, уровень сущностей.

Приложение. Краткий словарь терминов

Атрибут (от лат. attributio – приписывание, признак) – необходимое, существенное, неотъемлемое свойство предмета или явления.

Ключ сущности (англ. key of entity – ключевой атрибут) – атрибут, уникально определяющий сущность.

Компартментализация (от англ. compartment – отделение, отсек) – структурная область локализации того или иного вещества в клетке.

Модель сущность-связь (ER-модель) (англ. entity-relationship model, ERM) — модель данных, позволяющая описывать концептуальные схемы предметной области. ER-модель используется при высокоуровневом (концептуальном) проектировании баз данных. С её помощью можно выделить ключевые сущности и обозначить связи, которые могут устанавливаться между этими сущностями.

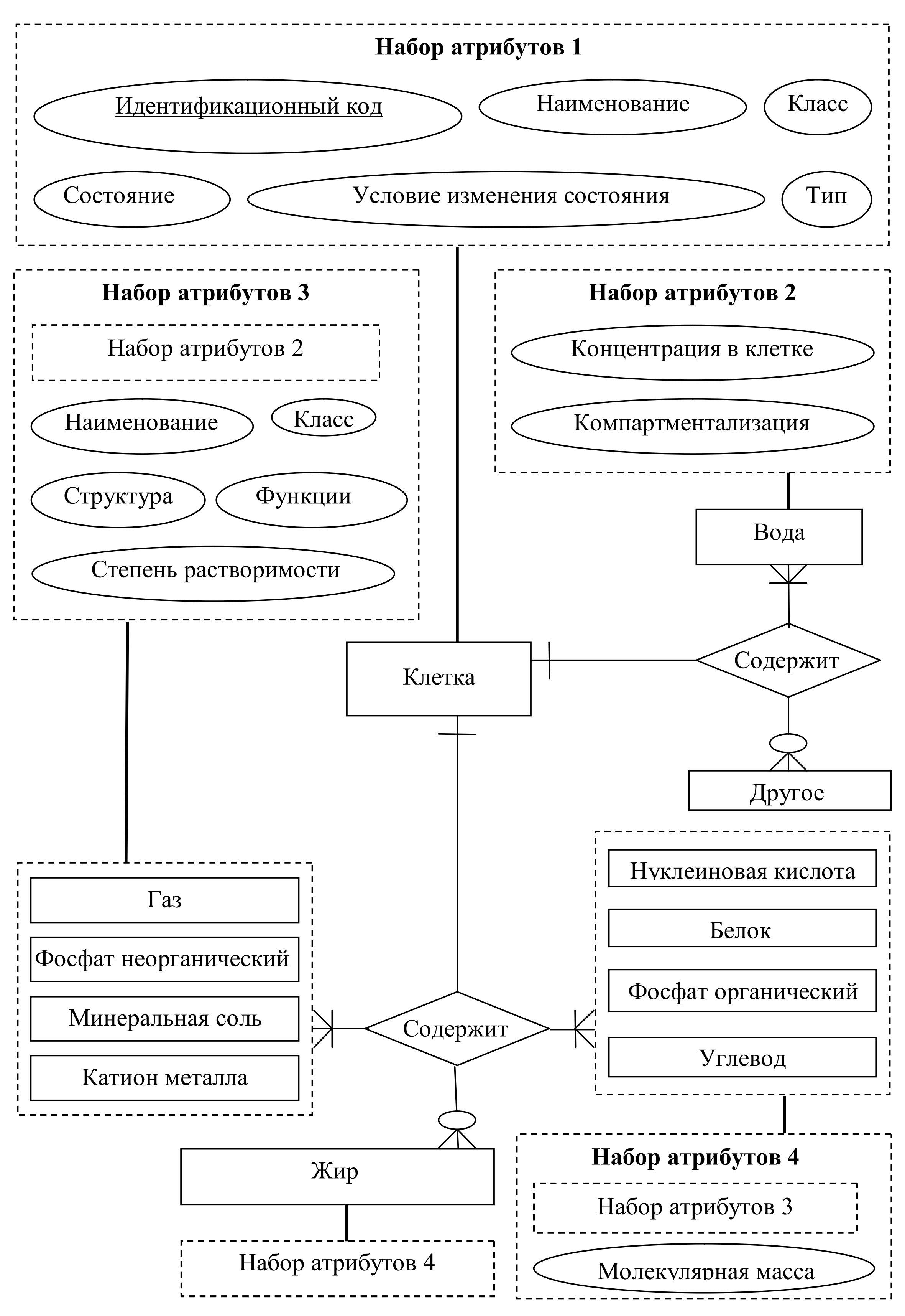


Рисунок 7 – ER-модель молекулярного состава клетки

Набор сущностей (англ. entity set) – множество сущностей одного типа, множество сущностей, обладающих одинаковыми свойствами.

Связь (англ. relationship) – ассоциация, установленная между несколькими сущностями.

Сущность (англ. entity) – объект, который может быть идентифицирован неким способом, отличающим его от других объектов. Примеры: конкретный человек, предприятие, событие и т.д.

Цитокомпьютер – компьютер, функционирование которого основано на алгоритмах клеточной жизнедеятельности.

Список литературы

1. Sterling T.I Think We Will Never Reach Zettaflops / Sterling T.I. [Электронный ресурс]. – Режим доступа: http://www.hpcwire.com/hpcwire/2012-05-07/thomas_sterling:_i_think_we_will_never_reach_zettaflops_.html.
2. Биокомпьютер [Электронный ресурс]. – Режим доступа: http://ru.wikipedia.org/wiki/Биокомпьютер.
3. DNA computing / [Błasiak J., Krasiński T., Popławski T., Sakowski S.] // Postepy Biochem.– 2011.– V. 57(1).– P. 13-23.
4. Терентьев А.А. Динамическая протеомика в моделировании живой клетки. Белок-белковые взаимодействия в моделировании живой клетки / А.А. Терентьев, Н.Т. Молдогазиева, К.В. Шайтан [Электронный ресурс]. – Режим доступа : http://www.inbi.ras.ru/ubkh/49/Terentiev.pdf.
5. Anfinsen C.B. Principles that Govern the Folding of Protein Chains / C.B. Anfinsen // Science. – 1973. – B. 4096. – T. 181. – C. 223-230.
6. Зеленков Ю.А. Введение в базы данных. Диаграмма «сущность-связь» / Зеленков Ю.А. [Электронный ресурс]. – Режим доступа: http://www.mstu.edu.ru/study/materials/zelenkov/ch_2_2.html.
7. Пушников А.Ю. Введение в системы управления базами данных. Элементы модели «сущность-связь» / Пушников А.Ю. [Электронный ресурс]. – Режим доступа: http://citforum.ru/database/dblearn/dblearn08.shtml.
8. In vitro selection of RNA-based irreversible inhibitors of human neutrophil elastase / [Smith D., Kirschenheuter G. P., Charlton J., Guidot D. M., Repine J.E.] // Chem. Biol. – 1995. – V.2. – P.741-750.
9. Famulok M. In vivo applied functional RNAs as tools in proteomics and genomics research / M. Famulok, S.Verma // Trends Biotechnol. – 2002. – V. 20. – P.462-466.
10. A novel RNA motif that binds efficiently and specifically to the Tat protein of HIV and inhibits the trans-activation by Tat of transcription in vitro and in vivo / [Yamamoto R., Katahira M., Nishikawa S., Baba T., Taira K., Kumar P.K.] // Genes Cells. – 2000. – V.5. – P.371-388.

References

1. Sterling T.: “I Think We Will Never Reach Zettaflops”. [Electronic resource]. – Access mode: http://hpc-mscman.blogspot.com/
2. Biokomputer [Electronic resource]. – Access mode: http://ru.wikipedia.org/wiki/Биокомпьютер.
3. Błasiak J., Krasiński T., Popławski T., Sakowski S. DNA computing // Postepy Biochem.– 2011.– V. 57(1).– P. 13– 23.
4. Terent'ev A. A., Moldogazieva N. T. , Shaitan K. V. Dynamic proteomics in modeling of a living cell. Protein-protein interactions in of a modeling of a living cell. [Electronic resource]. – Access mode: http://www.inbi.ras.ru/ubkh/49/Terentiev.pdf.
5. Anfinsen C. B. Principles that Govern the Folding of Protein Chains // Science. – 1973. – B. 4096. – T. 181. – C. 223–230.
6. Zelenkov Y. A. Introduction to Databases. The diagram “Entity-Relationship” [Electronic resource]. – Access mode: http://www.mstu.edu.ru/study/materials/zelenkov/ch_2_2.html.
7. Pushnikov A. Y. An introduction to the Database Management System. Elements of the model “Entity-Relationship” [Electronic resource]. – Access mode: http://citforum.ru/database/dblearn/dblearn08.shtml.
8. Smith D., Kirschenheuter G. P., Charlton J., Guidot D. M., Repine J. E. In vitro selection of RNA-based irreversible inhibitors of human neutrophil elastase // Chem. Biol. 1995. V.2. P.741–750.

9. Famulok M. , Verma S. In vivo applied functional RNAs as tools in proteomics and genomics research // Trends Biotechnol. 2002. V. 20. P.462-466.
10. Yamamoto R., Katahira M. , Nishikawa S., Baba T., Taira K., Kumar P. K. A novel RNA motif that binds efficiently and specifically to the Tat protein of HIV and inhibits the trans-activation by Tat of transcription in vitro and in vivo // Genes Cells. 2000. V.5. P.371–388.

RESUME

I.G. Gerasimov, S.V. Tereshchenko

Construction of Er-Model of the Molecular Composition of Cell As The Basis On Functioning Of Biocomputer (Cytocomputer)

Semiconductor technologies that stably develop a modern computer industry are now fast approaching to their physical and constructive limits.

Development of computers with the use of technology of silicic microchips can be limited due to their further miniaturization and growth of the exchange of information between the central processing unit and random access memory. Therefore becomes necessary create such computers, functioning of which is radically different from this technology.

An alternative to classical computers based on silicon microchips can be biocomputers. Biocomputer is a computer that contains biological components or functions as a living organism. In the latter case, in order to create a biocomputers is used such line of research as a molecular computing. As the computational elements are used proteins and nucleic acids, which react with each other. To simulate these interactions can be used ER-model.

The paper discusses the possibility for developing of a cytocomputer. Cytocomputer is a biocomputer which based on algorithms of cellular functioning and on available material base (chips). For this purpose was made the first attempt to construct an ER-model of the molecular structure of cells and were defined the relevant concepts.

Developed ER-model describes the molecular composition of cells. On the basis of this model and the selected groups of entities is possible to construct of the structural elements of cells which are necessary for modeling biocomputer based on algorithms cellular functioning.

Статья поступила в редакцию 27.12.2013.